



*"Record Protection in
an Uncertain World"*

WORLD CONFERENCE ON RECORDS AND GENEALOGICAL SEMINAR

Salt Lake City, Utah, U.S.A.
5-8 August 1969

COMPUTER INPUT AND OUTPUT TOOLS AND TECHNIQUES

Part I

The Input and Output Syndrome

By

Barbara G. Fisher

COMPUTER INPUT AND OUTPUT TOOLS AND TECHNIQUES

Part I

The Input and Output Syndrome

By

Barbara G. Fisher

A syndrome is a given set of symptoms characteristic of a particular disease. While the input-output complex, which I describe as the I/O syndrome, is not indicative of disease, input and output content, methods, and machines sometimes lead to a sense of frustrating malaise. Choice of the right or the best machine can be an irritating factor in the development of automated systems because input and output content and form are compounded of many cells of logic, each of which must be diagnosed and shaped to machine capability. First of all, however, we must be physicians to our own intelligence, and ask ourselves what we want to automate and why.

The prescription for any automated or mechanized system calls for a combining formula of readability, compatability, and retrievability. In short, an integrated processing system must respect data requirements, machine components, their capabilities, and potential user needs. The effectiveness of any machine system, however, first requires that exercise of judgment known in computer jargon as feasibility. A feasible project is simply one that is practical in terms of data preparation, use of available hardware, and over-all cost.

Feasibility studies of both manual and automatic information systems are basic diagnostic steps that we must take before any information or management system can be committed or converted to computer control. Computers can do no more than they are instructed to do. We must prepare those instructions on a common sense basis of intelligent analysis of what we want the machine to do and why.

Before we decide what data to make machine-readable, what programs will be clearly compatible with the machines available, or how we would like to retrieve stored data, we must grapple with stern realities-hardware limitations, operational costs, man-hours and machine time, expected results, and basic purpose. In addition, distinctions must be made between operations that can best be handled by unit record or data processing equipment, and those operations that require total computer system design.

A computerized system is practical only after a hard look at data accumulation problems, projected benefits, and at the underlying principles of computer technology. Poor or useless data will not assume excellence or usefulness simply by putting it on magnetic tape or punched

cards. Meaningless manipulations of data will not suddenly assume meaningful form simply because the manipulations have taken place at great speed. If input on a sophisticated IBM 2741 terminal is made up of "garbage," this is exactly what the terminal will transmit and exactly what the printer will print--"garbage." Only transmigrations, not transmutations, take place in the computer.

The computer configuration--the machine battery that includes memory units, data processing machines, tape and disk drives, terminals--is increasingly becoming a communications rather than a processing configuration. Engineering design is turning more and more to computer storage and on-line access, and away from simple one-time or batch processing. Archivists and genealogists are also increasingly concerned with accumulation and storage of great masses of reference information, centralized in one file but accessible to single inquiry as well as to retrieval for future selected assembly of data for publication.

Evolution of input equipment from the keypunch machine to direct access or on-line terminals illustrates a continuing effort to narrow the gap between man-generated source data and computer-stored information, and to reduce the time between conversion of information to machine-readable form and to the ultimate production of output.

Choice of input equipment depends primarily on the need for rapid source data conversion, massive accumulation of data, and on the continuous necessity to update stored information. The period of wide choice in input machine selection may be drawing to a close, however, because more and more input devices are falling victim to the peculiarly American idea of obsolescence. The characteristic American thought that old models should be replaced or traded in when new models appear is nowhere more vividly demonstrated than in the rapid evolution of IBM computers-- from the 1400 series to the 360 series. Indeed, the computer industry has been so preoccupied with technical advance that it has seldom been conscious of the human need to progress by gradual stages. Increasing competition in computer manufacturing is producing a great proliferation of equipment and will undoubtedly lead to substantial cost reduction, but it is also adding to the momentum of obsolescence. If the rate of hardware competition continues to accelerate, paradoxically enough there will soon be less and less choice of input equipment, as manufacturers tend to conform to best-selling models.

A brief history of input equipment demonstrates that computer design has been directed to narrowing the gap between raw source data and computer-stored information. Initial electric or electronic systems, still less than 50 years old, required:

1. Manual transcription of data from source documents to data sheets.
2. Key punching of data from the data sheets.
3. Verification and sorting.
4. Printout on a tabulator-printer (a line printer or listing machine).

The second machine effort, designed to reduce the processing time gap, was the continuous roll paper tape punch. The Friden Flexowriter is representative of this kind of input device. This system requires:

1. Manual transcription of data on data sheets.
2. Simultaneous punching onto paper tape.
3. Conversion of paper tape to magnetic tape (this generally requires special converter hardware).
4. Conversion of paper tape to cards to magnetic tape (a required system when no converter hardware is available).

An input device which permits input data to be transcribed directly onto mini-reels of magnetic tape compatible with the bit, byte and track components of computer tape drives is exemplified by Mohawk 1101 magnetic data tape recorder.

The optical scanner, developed about the same time as the paper tape input machine, was designed to transfer data directly from source documents to computer magnetic tape. The process greatly reduces the time required for data conversion although formatting of input data is of necessity strictly standardized. A new direction in input devices, still in developmental stages, is the laser beam sensor.

The most rapid input and access device currently in use is the terminal or on-line keyboard-an electronic keyboard linked to the master computer by cable. The terminal can carry messages and data directly into the computer, can interrogate stored data, and on instruction can feed back rapid response to inquiry. The terminal input device provides almost instantaneous communication between the computer input, data storage, and output sequences.

All five input systems are currently in use, sometimes in the same computer center. Each device has distinctive job advantages. The card keypunch system is an excellent teaching tool for newcomers to computer processing. It is a visual aid to understanding the logic of input design and the logic of data definition. It also teaches principles of automatic control of data input through program drum control cards, and the use of automatic feed, duplication, and skip controls. The card punch system permits total visual and intellectual operator control of input data. Associated unit record equipment-verifiers, sorters, collators and interpreters-also provide insight into error checking procedures, coding, and data manipulation. The card keypunch machine is the most widespread of all input machines because rental is reasonably moderate, and anyone who can type can easily learn to keypunch.

A by-product of card punching systems is the accumulation of a working card file during the preliminary stages of testing or experimental programs. One of the advantages of the card

system is the multiple usefulness of cards to many users. In addition, card files may be processed in either the computer or unit record machines. Most other input devices are restricted to machine specialists or operators, are generally housed in a closed shop, and input data can be interpreted only by the computer. The disadvantages of punched cards are the necessity for special storage and filing of card decks, the possibility of card damage by tearing or warping, and the confining concept of only 80 columns for the transcription of data.

While paper tape and magnetic tape input devices permit the mass production of converted data with automatic control of format errors, paper tape processing by the computer is relatively slow. The computer can interpret only 60 tape-recorded characters per second whereas punched cards can be interpreted at the approximate rate of 1280 characters per second. Many tape input devices have presented substantial machine compatibility problems and many computer centers cannot handle a diversity of input forms economically. The computer and its components operate on integrated and exacting timing schedules. Irregular speeds at any phase of the processing--program assembly, data interpretation, core storage, tape-to-tape search and retrieval, and printing--can badly backlog, delay, or disrupt the system. Paper tape machines are particularly useful, however, for duplicate copy making, for form letters and for administrative reports. They have demonstrated their usefulness consistently in the data processing activities of the American Genealogical Society.

But input is only the ingress phase of the computer processing sequence. One of the major dilemmas of the I/O syndrome at the output phase is the unmanageable masses of paper printout that can result from a total output program. Endless paper output is one aspect of the information explosion--a computer-generated information explosion, about which little comment has been made. A computer store of 5,000,000 characters, for example, could produce about 100,000 lines of text that could be printed in about 1 1/2 hours. At that rate, a printout of stored data might run to over 2,000 pages. In the foreseeable future low cost mass storage will accommodate 600,000,000 characters. This could conceivably result in a printout of 12,000,000 lines of text approximating some 200,000 pages of print. It is awesome to realize that a text this large could be processed by the high speed IBM 1403 printer in about 3 hours.

The high speed printer is a necessity in a well-integrated computer layout where many small as well as large projects are processed. Integrated also with a terminal access system, the high speed printer can produce selective hard copy output at the same time the computer is responding directly to inquiry. The printer is designed to handle many jobs rapidly as well as to produce extensive printouts.

Resolution of the phenomena of too much information requires the same logical tests of practicality as those that are applied to the determination of input. In fact, output should be determined even prior to final definition of input. The consequences of any projected system are as important as the sequences. In most present computer installations and in current software packages, output must of necessity be selective.

As librarians, bibliographers, archivists, genealogists, we can estimate reasonably well, on the basis of established criteria and accumulated experience, what elements of data in a master record are most often needed. Selective retrieval and print programs can be designed to produce both economical and pertinent professional indexes, listings or guides.

Increasing momentum toward on-line computer services, and wide dispersal of terminal keyboards, however, will soon shift the initiative for meeting real needs for stored information from the librarian to the user. The computer processing sequence of the future will expand from an input-storage-output cycle to an input-storage-use record and output or response cycle. Just as oral magnetic tape now preserves random or informal responses and opinions, computer recording media can preserve a record of the uses made of stored information.

By-products of this enlarged processing cycle could include computer assembled thesauri of search of key terms, subject bibliographies, and even work-in-progress or user-trend reports.

The computerized data bank of the future will yield to the on-line inquirer not only the data he needs and wants; it will also record his method of search, accumulate search patterns, and will reduce demand for massive printouts. Output will more and more become user-generated and optionalized, tailored to meet specific user need.

The emphasis in input machine design is on reducing the necessity for preselection and pre-editing of information. The emphasis in output is on increasing the immediacy of the dialogue between inquirer and computer-stored information. The trend in computer technology is toward automata and self-generating systems. To prepare for this text realization of computer potential the newcomer to automation should study the history of machine developments from the abacus to the IBM 360, from the typewriter to the terminal keyboard. While machine concepts of data recording and manipulation do not provide insight into the meaning of information, they do provide clues to methods of human information gathering and communication.

Input and output concepts, methods and machines need not be a syndrome marked by irritation and malaise. Built on the sound diagnosis of human and machine intelligence systems, the I/O complex should represent a synthesis of cause and effect, and a unity of plan and product.